

WHAT IS CLAIMED IS:

1. A method of maintaining a multidimensional histogram for a data array having a data array size, the method having a processing time substantially less than proportional to the data array size, the method comprising:

receiving a data update that indicates a change to data in the data array;

with the data update, updating an intermediate data structure having a size substantially smaller than the data array size, so that the updated intermediate data structure remains an at-least-approximate representation of the data in the data array as changed by the data update;

collecting a number of substantially-largest-coefficient linear combinations of then-current data, the number being small compared with the data array size; and

forming the multidimensional histogram as a histogram to an intermediate data array re-synthesized from the collected linear combinations.

2. The method of Claim 1, wherein:

the step of forming the multidimensional histogram is executed unconditionally in response to reception of the data update.

3. The method of Claim 1, wherein:

the method further comprises determining whether or not a multidimensional histogram should be formed in response to reception of the data update; and

the step of forming the multidimensional histogram is executed conditionally, in response to an outcome of the determining step that the multidimensional histogram should be formed in response to the reception of the data update.

4. The method of Claim 1, wherein the intermediate data structure constitutes a histogram sketch of the data that includes one or both of:

an identification sketch of linear combinations of the data, the identification sketch being configured to identify, with limited error, linear combinations whose coefficients are large; and

a count sketch of linear combinations of data, the count sketch being configured to approximate the coefficients of linear combinations.

5. The method of Claim 1, wherein:

the linear combinations are tensor products of Haar wavelets (TPHWs) of data in the data array.

6. The method of Claim 1, further comprising:

reducing the collection of linear combinations before forming the multidimensional histogram, by discarding a linear combination according to a criterion belonging to a group of criteria including:

a) a square of a largest coefficient of a discarded linear combination is small compared with a sum of squares of coefficients of all discarded linear combinations;

b) a square of a largest coefficient of a discarded linear combination is small compared with a sum of squares of a difference between a then-current data and an array represented by a collection of retained linear combinations; and

c) a square of a largest coefficient of a discarded linear combination is small compared with a sum of squares of a difference between a then-current data and an array represented by a collection of retained linear combinations, the sum of squares being approximated using a sum-of-squares sketch.

7. The method of Claim 1, wherein the step of updating the intermediate data structure includes:

combining plural pre-intermediate data structures, each of which may not itself be a faithful representation of the data in the data array.

8. A method of preparing a multidimensional histogram for a data array, the data array characterized by a data array size and including data, the method having an execution time proportional to the data array size and using an amount of storage space substantially smaller than the data array size, the method comprising:

receiving data from the data array;

transforming the data into linear combinations of data items, or of approximations of the linear combinations of data items;

collecting a moderate number of substantially-largest-coefficient linear combinations of the data; and

forming the multidimensional histogram as a histogram to the collected linear combinations.

9. The method of Claim 8, wherein the data receiving step includes:

reading the data in a reading sequence such that at least some of the data that is close together in the reading sequence is close together in a multidimensional grid, so as to facilitate the transforming step.

10. The method of Claim 8, wherein:

the linear combinations are tensor products of Haar wavelets (TPHWs) of data in the data array.

11. The method of Claim 8, further comprising:

reducing the collection of linear combinations before forming the multidimensional histogram, by discarding a linear combination according to a criterion belonging to a group of criteria including:

a) a square of a largest coefficient of a discarded linear combination is small compared with a sum of squares of coefficients of all discarded linear combinations;

- b) a square of a largest coefficient of a discarded linear combination is small compared with a sum of squares of a difference between a then-current data and an array represented by a collection of retained linear combinations; and
- c) a square of a largest coefficient of a discarded linear combination is small compared with a sum of squares of a difference between a then-current data and an array represented by a collection of retained linear combinations, the sum of squares being approximated using a sum-of-squares sketch.

12. A computer program product including computer executable code or computer executable instructions that, when executed, causes a computer system including at least one processing device to perform the method of Claim 1.

13. A computer program product including computer executable code or computer executable instructions that, when executed, causes a computer system including at least one processing device to perform the method of Claim 4.

14. A computer program product including computer executable code or computer executable instructions that, when executed, causes a computer system including at least one processing device to perform the method of Claim 8.

15. A computer program product including computer executable code or computer executable instructions that, when executed, causes a computer system including at least one processing device to perform the method of Claim 9.

16. A system configured to perform the method of Claim 1.

17. A system configured to perform the method of Claim 4.

18. A system configured to perform the method of Claim 8.

19. A system configured to perform the method of Claim 9.
20. A system configured to perform the method of Claim 10.